



# wwPDB X-ray Structure Validation Summary Report

Aug 7, 2014 – 07:03 PM EDT

PDB ID : 4UNZ  
Title : Structure of the A\_Equine\_Newmarket\_2\_93H3 haemagglutinin in complex with 6SO4-Sialyl Lewis X  
Authors : Vachieri, S.G.; Collins, P.J.; Haire, L.F.; Ogrodowicz, R.W.; Martin, S.R.; Walker, P.A.; Xiong, X.; Gamblin, S.J.; Skehel, J.J.  
Deposited on : 2014-05-31  
Resolution : 2.90 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

---

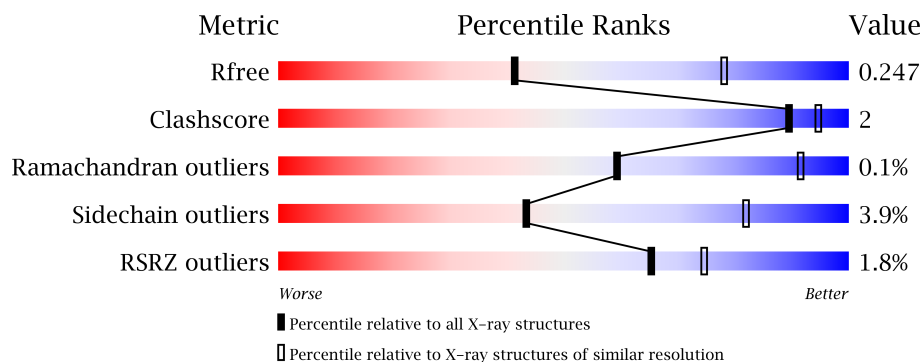
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.16 November 2013  
Xtriage (Phenix) : dev-1439  
EDS : stable23489  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable23489

# 1 Overall quality at a glance

The reported resolution of this entry is 2.90 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	66092	1053 (2.90-2.90)
Clashscore	79885	1326 (2.90-2.90)
Ramachandran outliers	78287	1290 (2.90-2.90)
Sidechain outliers	78261	1292 (2.90-2.90)
RSRZ outliers	66119	1054 (2.90-2.90)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	330	
1	C	330	
1	E	330	
2	B	173	
2	D	173	
2	F	173	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
3	NAG	A	421	-	X
3	NAG	A	601	-	X

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Geometry	Electron density
3	NAG	C	471	-	X
3	NAG	E	411	-	X
3	NAG	F	201	-	X

## 2 Entry composition

There are 11 unique types of molecules in this entry. The entry contains 12397 atoms, of which 0 are hydrogen and 0 are deuterium.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called HAY SUBUNIT OF HAEMAGGLUTININ.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	317	Total	C	N	O	S	0	0	0
			2463	1540	433	476	14			
1	C	330	Total	C	N	O	S	0	0	0
			2564	1599	454	497	14			
1	E	318	Total	C	N	O	S	0	0	0
			2447	1528	431	474	14			

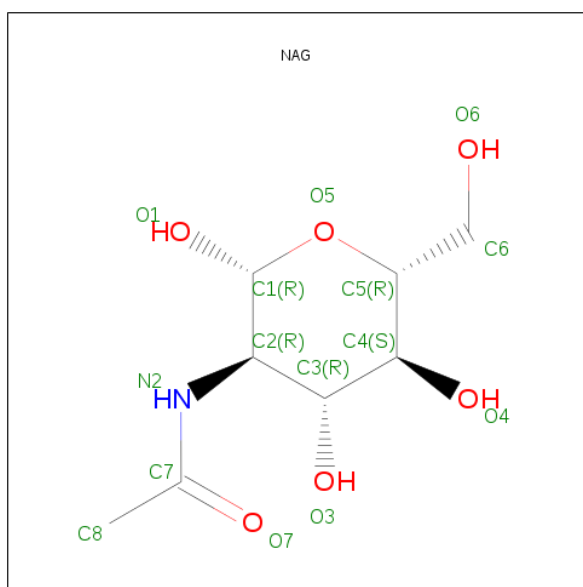
There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	PRO	-	EXPRESSION TAG	UNP Q82847
A	1	ASP	-	EXPRESSION TAG	UNP Q82847
C	0	PRO	-	EXPRESSION TAG	UNP Q82847
C	1	ASP	-	EXPRESSION TAG	UNP Q82847
E	0	PRO	-	EXPRESSION TAG	UNP Q82847
E	1	ASP	-	EXPRESSION TAG	UNP Q82847

- Molecule 2 is a protein called H3 HAEMAGGLUTININ HA2 CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	173	Total	C	N	O	S	0	0	0
			1404	876	246	276	6			
2	D	172	Total	C	N	O	S	0	0	0
			1396	870	245	275	6			
2	F	172	Total	C	N	O	S	0	0	0
			1388	864	243	275	6			

- Molecule 3 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	C	1	Total	C	N	O	0	0
			14	8	1	5		
3	E	1	Total	C	N	O	0	0
			14	8	1	5		
3	F	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 4 is a polymer of unknown type called SUGAR (4-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	4	Total	C	N	O	0	0
			50	28	2	20		

- Molecule 5 is a polymer of unknown type called SUGAR (4-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	4	Total	C	N	O	S	0	0
			60	31	2	26	1		
5	C	4	Total	C	N	O	S	0	0
			60	31	2	26	1		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	E	4	Total	C	N	O	S	
			60	31	2	26	1	
							0	0

- Molecule 6 is a polymer of unknown type called SUGAR (3-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	B	3	Total	C	N	O		
			38	22	2	14		
							0	0

- Molecule 7 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
7	C	2	Total	C	N	O		
			28	16	2	10		
							0	0
7	C	2	Total	C	N	O		
			28	16	2	10		
							0	0
7	C	2	Total	C	N	O		
			28	16	2	10		
							0	0
7	C	2	Total	C	N	O		
			28	16	2	10		
							0	0
7	E	2	Total	C	N	O		
			28	16	2	10		
							0	0
7	E	2	Total	C	N	O		
			28	16	2	10		
							0	0

- Molecule 8 is a polymer of unknown type called SUGAR (3-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
8	C	3	Total	C	N	O		
			39	22	2	15		
							0	0
8	E	3	Total	C	N	O		
			39	22	2	15		
							0	0

- Molecule 9 is a polymer of unknown type called SUGAR (6-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
9	C	6	Total	C	N	O		
			71	40	2	29		
							0	0

- Molecule 10 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
10	D	2	Total	C	N	O	0	0
			24	14	1	9		

- Molecule 11 is water.

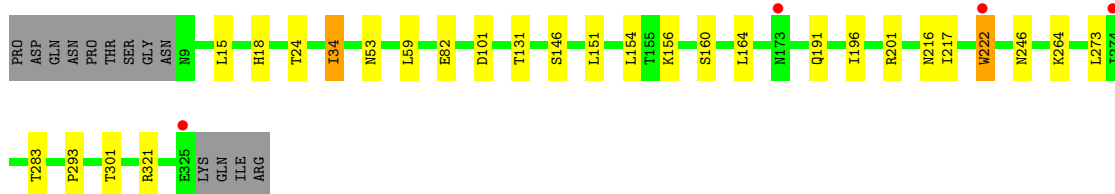
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
11	A	12	Total	O	0	0
			12	12		
11	B	5	Total	O	0	0
			5	5		
11	C	12	Total	O	0	0
			12	12		
11	D	8	Total	O	0	0
			8	8		
11	E	3	Total	O	0	0
			3	3		
11	F	2	Total	O	0	0
			2	2		

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: HAY SUBUNIT OF HAEMAGGLUTININ

Chain A: 



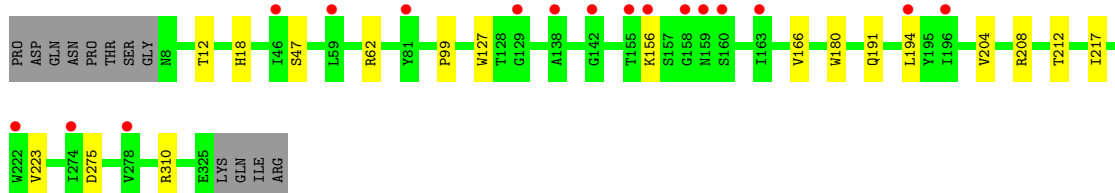
- Molecule 1: HAY SUBUNIT OF HAEMAGGLUTININ

Chain C: 



- Molecule 1: HAY SUBUNIT OF HAEMAGGLUTININ

Chain E: 



- Molecule 2: H3 HAEMAGGLUTININ HA2 CHAIN

Chain B: 



- Molecule 2: H3 HAEMAGGLUTININ HA2 CHAIN

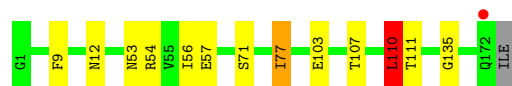
Chain D: 





- Molecule 2: H3 HAEMAGGLUTININ HA2 CHAIN

Chain F: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	99.55Å 125.75Å 165.11Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	100.04 – 2.90 47.66 – 2.90	Depositor EDS
% Data completeness (in resolution range)	98.2 (100.04-2.90) 98.2 (47.66-2.90)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.16 (at 2.91Å)	Xtriage
Refinement program	REFMAC 5.8.0069	Depositor
R, $R_{free}$	0.198 , 0.248 0.203 , 0.247	Depositor DCC
$R_{free}$ test set	2333 reflections (5.37%)	DCC
Wilson B-factor (Å <sup>2</sup> )	66.3	Xtriage
Anisotropy	0.544	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 37.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	1 of 45818 reflections (0.002%)	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	12397	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	75.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.51% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: BMA, NAG, SIA, GAL, FUC, NGS, MAN

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	# Z  >5	RMSZ	# Z  >5
1	A	0.30	0/2516	0.54	0/3419
1	C	0.30	0/2619	0.55	0/3558
1	E	0.30	0/2500	0.50	0/3404
2	B	0.34	0/1429	0.57	1/1921 (0.1%)
2	D	0.33	0/1421	0.57	1/1910 (0.1%)
2	F	0.31	0/1413	0.51	1/1902 (0.1%)
All	All	0.31	0/11898	0.54	3/16114 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
5	A	1	0
5	C	1	0
5	E	3	0
6	B	1	0
7	C	1	0
9	C	1	0
10	D	1	0
All	All	9	0

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	110	LEU	CA-CB-CG	6.87	131.09	115.30
2	B	110	LEU	CA-CB-CG	6.77	130.87	115.30
2	D	110	LEU	CA-CB-CG	6.34	129.87	115.30

5 of 9 chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
5	A	704	FUC	C1
6	B	200	FUC	C1
9	C	450	FUC	C1
7	C	601	NAG	C1
5	C	704	FUC	C1

There are no planarity outliers.

## 5.2 Close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2463	0	2398	13	0
1	C	2564	0	2496	10	0
1	E	2447	0	2352	5	0
2	B	1404	0	1328	11	0
2	D	1396	0	1317	6	0
2	F	1388	0	1295	9	0
3	A	42	0	39	2	0
3	C	14	0	13	0	0
3	E	14	0	13	0	0
3	F	14	0	13	0	0
4	A	50	0	43	0	0
5	A	60	0	49	1	0
5	C	60	0	49	1	0
5	E	60	0	49	0	0
6	B	38	0	34	1	0
7	C	112	0	100	1	0
7	E	56	0	50	0	0
8	C	39	0	34	0	0
8	E	39	0	34	2	0
9	C	71	0	61	2	0
10	D	24	0	22	0	0
11	A	12	0	0	0	0
11	B	5	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
11	C	12	0	0	0	0
11	D	8	0	0	0	0
11	E	3	0	0	0	0
11	F	2	0	0	0	0
All	All	12397	0	11789	52	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 2.

The worst 5 of 52 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
2:B:107:THR:HA	2:B:110:LEU:HD13	1.78	0.65
2:F:56:ILE:HG22	2:F:57:GLU:HG2	1.82	0.61
2:B:76:ARG:NE	2:F:77:ILE:HD11	2.17	0.60
1:A:15:LEU:HD22	2:B:119:PHE:HA	1.84	0.59
2:B:77:ILE:HD11	2:D:76:ARG:NE	2.18	0.59

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	315/330 (96%)	296 (94%)	19 (6%)	0	100	100
1	C	328/330 (99%)	309 (94%)	19 (6%)	0	100	100
1	E	316/330 (96%)	288 (91%)	27 (8%)	1 (0%)	50	85
2	B	171/173 (99%)	157 (92%)	13 (8%)	1 (1%)	33	76
2	D	170/173 (98%)	161 (95%)	9 (5%)	0	100	100
2	F	170/173 (98%)	161 (95%)	9 (5%)	0	100	100
All	All	1470/1509 (97%)	1372 (93%)	96 (6%)	2 (0%)	59	91

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	57	GLU
1	E	62	ARG

### 5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution. The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	277/290 (96%)	264 (95%)	13 (5%)	36	75
1	C	289/290 (100%)	281 (97%)	8 (3%)	56	90
1	E	272/290 (94%)	264 (97%)	8 (3%)	55	89
2	B	145/145 (100%)	139 (96%)	6 (4%)	41	80
2	D	144/145 (99%)	135 (94%)	9 (6%)	25	60
2	F	142/145 (98%)	137 (96%)	5 (4%)	48	85
All	All	1269/1305 (97%)	1220 (96%)	49 (4%)	43	82

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	156	LYS
2	D	11	GLU
2	F	53	ASN
1	C	222	TRP
2	D	12	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	211	GLN
2	F	53	ASN
1	C	211	GLN
1	A	53	ASN
2	B	60	ASN

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

45 carbohydrates are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# $ Z  > 2$	Counts	RMSZ	# $ Z  > 2$
4	NAG	A	431	1,4	12,14,15	0.76	1 (8%)	15,19,21	1.43	1 (6%)
4	NAG	A	432	4	12,14,15	0.73	1 (8%)	15,19,21	0.99	0
4	BMA	A	433	4	10,11,12	0.65	0	11,15,17	1.23	2 (18%)
4	MAN	A	437	4	10,11,12	0.84	1 (10%)	11,15,17	0.83	0
5	SIA	A	701	5	20,20,21	3.89	1 (5%)	23,28,31	2.92	4 (17%)
5	GAL	A	702	5	10,11,12	0.96	1 (10%)	11,15,17	1.21	1 (9%)
5	NGS	A	703	5	19,19,19	0.77	1 (5%)	28,28,28	0.99	2 (7%)
5	FUC	A	704	5	9,10,11	0.87	0	10,14,16	1.72	3 (30%)
6	FUC	B	200	6	9,10,11	0.94	1 (11%)	10,14,16	0.75	0
6	NAG	B	201	2,6	12,14,15	0.79	1 (8%)	15,19,21	0.80	0
6	NAG	B	202	6	12,14,15	0.69	1 (8%)	15,19,21	1.06	2 (13%)
7	NAG	C	411	1,7	12,14,15	0.66	0	15,19,21	1.11	1 (6%)
7	NAG	C	412	7	12,14,15	0.76	1 (8%)	15,19,21	0.96	2 (13%)
7	NAG	C	421	1,7	12,14,15	0.71	1 (8%)	15,19,21	1.34	2 (13%)
7	NAG	C	422	7	12,14,15	0.75	1 (8%)	15,19,21	0.88	1 (6%)
7	NAG	C	431	1,7	12,14,15	0.74	1 (8%)	15,19,21	1.07	1 (6%)
7	NAG	C	432	7	12,14,15	0.61	0	15,19,21	1.11	1 (6%)
8	NAG	C	441	1,8	12,14,15	0.80	1 (8%)	15,19,21	1.05	1 (6%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
8	NAG	C	442	8	12,14,15	0.74	1 (8%)	15,19,21	0.75	0
8	BMA	C	443	8	10,11,12	0.55	0	11,15,17	0.76	1 (9%)
9	FUC	C	450	9	9,10,11	1.17	1 (11%)	10,14,16	1.22	1 (10%)
9	NAG	C	451	1,9	12,14,15	1.04	1 (8%)	15,19,21	1.14	1 (6%)
9	NAG	C	452	9	12,14,15	0.86	1 (8%)	15,19,21	1.35	3 (20%)
9	BMA	C	453	9	10,11,12	0.67	0	11,15,17	1.41	1 (9%)
9	MAN	C	454	9	10,11,12	0.77	0	11,15,17	1.03	1 (9%)
9	MAN	C	457	9	10,11,12	0.83	1 (10%)	11,15,17	1.15	2 (18%)
7	NAG	C	601	1,7	12,14,15	0.58	0	15,19,21	1.58	1 (6%)
7	NAG	C	602	7	12,14,15	0.62	0	15,19,21	1.47	3 (20%)
5	SIA	C	701	5	20,20,21	4.01	1 (5%)	23,28,31	2.82	3 (13%)
5	GAL	C	702	5	10,11,12	0.91	1 (10%)	11,15,17	1.24	1 (9%)
5	NGS	C	703	5	19,19,19	0.83	1 (5%)	28,28,28	1.42	4 (14%)
5	FUC	C	704	5	9,10,11	0.97	1 (11%)	10,14,16	0.81	0
10	FUC	D	200	10	9,10,11	0.93	1 (11%)	10,14,16	1.01	1 (10%)
10	NAG	D	201	10,2	12,14,15	0.57	0	15,19,21	1.41	3 (20%)
8	NAG	E	431	1,8	12,14,15	0.77	1 (8%)	15,19,21	0.87	1 (6%)
8	NAG	E	432	8	12,14,15	0.68	0	15,19,21	2.03	4 (26%)
8	BMA	E	433	8	10,11,12	0.65	0	11,15,17	0.73	0
7	NAG	E	441	1,7	12,14,15	0.79	1 (8%)	15,19,21	1.42	3 (20%)
7	NAG	E	442	7	12,14,15	0.69	1 (8%)	15,19,21	0.74	0
7	NAG	E	601	1,7	12,14,15	0.79	1 (8%)	15,19,21	1.66	5 (33%)
7	NAG	E	602	7	12,14,15	0.78	0	15,19,21	2.08	6 (40%)
5	SIA	E	701	5	20,20,21	4.07	1 (5%)	23,28,31	2.57	4 (17%)
5	GAL	E	702	5	10,11,12	0.83	1 (10%)	11,15,17	0.51	0
5	NGS	E	703	5	19,19,19	0.78	1 (5%)	28,28,28	1.13	2 (7%)
5	FUC	E	704	5	9,10,11	0.85	0	10,14,16	1.86	2 (20%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	NAG	A	431	1,4	-	0/6/23/26	0/1/1/1
4	NAG	A	432	4	-	0/6/23/26	0/1/1/1
4	BMA	A	433	4	-	0/2/19/22	0/1/1/1

Continued on next page...



*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	MAN	A	437	4	-	0/2/19/22	0/1/1/1
5	SIA	A	701	5	-	0/15/34/38	0/1/1/1
5	GAL	A	702	5	-	0/2/19/22	0/1/1/1
5	NGS	A	703	5	-	0/10/30/30	0/1/1/1
5	FUC	A	704	5	1/1/4/5	0/0/17/20	0/1/1/1
6	FUC	B	200	6	1/1/4/5	0/0/17/20	0/1/1/1
6	NAG	B	201	2,6	-	0/6/23/26	0/1/1/1
6	NAG	B	202	6	-	0/6/23/26	0/1/1/1
7	NAG	C	411	1,7	-	0/6/23/26	0/1/1/1
7	NAG	C	412	7	-	0/6/23/26	0/1/1/1
7	NAG	C	421	1,7	-	0/6/23/26	0/1/1/1
7	NAG	C	422	7	-	0/6/23/26	0/1/1/1
7	NAG	C	431	1,7	-	0/6/23/26	0/1/1/1
7	NAG	C	432	7	-	0/6/23/26	0/1/1/1
8	NAG	C	441	1,8	-	0/6/23/26	0/1/1/1
8	NAG	C	442	8	-	0/6/23/26	0/1/1/1
8	BMA	C	443	8	-	0/2/19/22	0/1/1/1
9	FUC	C	450	9	1/1/4/5	0/0/17/20	0/1/1/1
9	NAG	C	451	1,9	-	0/6/23/26	0/1/1/1
9	NAG	C	452	9	-	0/6/23/26	0/1/1/1
9	BMA	C	453	9	-	0/2/19/22	0/1/1/1
9	MAN	C	454	9	-	0/2/19/22	0/1/1/1
9	MAN	C	457	9	-	0/2/19/22	0/1/1/1
7	NAG	C	601	1,7	1/1/5/7	0/6/23/26	0/1/1/1
7	NAG	C	602	7	-	0/6/23/26	0/1/1/1
5	SIA	C	701	5	-	0/15/34/38	0/1/1/1
5	GAL	C	702	5	-	0/2/19/22	0/1/1/1
5	NGS	C	703	5	-	0/10/30/30	0/1/1/1
5	FUC	C	704	5	1/1/4/5	0/0/17/20	0/1/1/1
10	FUC	D	200	10	1/1/4/5	0/0/17/20	0/1/1/1
10	NAG	D	201	10,2	-	0/6/23/26	0/1/1/1
8	NAG	E	431	1,8	-	0/6/23/26	0/1/1/1
8	NAG	E	432	8	-	0/6/23/26	0/1/1/1
8	BMA	E	433	8	-	0/2/19/22	0/1/1/1
7	NAG	E	441	1,7	-	0/6/23/26	0/1/1/1
7	NAG	E	442	7	-	0/6/23/26	0/1/1/1
7	NAG	E	601	1,7	-	0/6/23/26	0/1/1/1
7	NAG	E	602	7	-	0/6/23/26	0/1/1/1
5	SIA	E	701	5	1/1/8/9	0/15/34/38	0/1/1/1
5	GAL	E	702	5	1/1/4/5	0/2/19/22	0/1/1/1
5	NGS	E	703	5	-	0/10/30/30	0/1/1/1
5	FUC	E	704	5	1/1/4/5	0/0/17/20	0/1/1/1

The worst 5 of 31 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	E	701	SIA	O6-C2	17.99	1.44	1.28
5	C	701	SIA	O6-C2	17.69	1.44	1.28
5	A	701	SIA	O6-C2	17.16	1.44	1.28
9	C	451	NAG	O5-C5	-3.18	1.40	1.45
9	C	450	FUC	O5-C5	-2.86	1.40	1.45

The worst 5 of 77 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	701	SIA	O6-C2-C3	-12.90	108.05	124.91
5	C	701	SIA	O6-C2-C3	-12.39	108.72	124.91
5	E	701	SIA	O6-C2-C3	-10.84	110.74	124.91
8	E	432	NAG	O5-C5-C6	5.03	112.26	106.98
4	A	431	NAG	O5-C5-C6	4.64	111.85	106.98

5 of 9 chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
5	A	704	FUC	C1
10	D	200	FUC	C1
5	C	704	FUC	C1
5	E	704	FUC	C1
9	C	450	FUC	C1

There are no torsion outliers.

There are no ring outliers.

## 5.6 Ligand geometry ⓘ

6 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	NAG	A	421	1	12,14,15	0.77	1 (8%)	15,19,21	1.83	5 (33%)
3	NAG	A	441	1	12,14,15	0.68	1 (8%)	15,19,21	1.62	5 (33%)
3	NAG	A	601	1	12,14,15	0.69	1 (8%)	15,19,21	1.30	1 (6%)
3	NAG	C	471	1	12,14,15	0.65	0	15,19,21	0.90	1 (6%)
3	NAG	E	411	1	12,14,15	0.70	1 (8%)	15,19,21	0.69	0
3	NAG	F	201	2	12,14,15	0.77	1 (8%)	15,19,21	0.89	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	NAG	A	421	1	-	0/6/23/26	0/1/1/1
3	NAG	A	441	1	-	0/6/23/26	0/1/1/1
3	NAG	A	601	1	-	0/6/23/26	0/1/1/1
3	NAG	C	471	1	1/1/5/7	0/6/23/26	0/1/1/1
3	NAG	E	411	1	-	0/6/23/26	0/1/1/1
3	NAG	F	201	2	-	0/6/23/26	0/1/1/1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	F	201	NAG	O5-C5	-2.40	1.41	1.45
3	E	411	NAG	O5-C5	-2.15	1.42	1.45
3	A	421	NAG	O5-C5	-2.09	1.42	1.45
3	A	441	NAG	O5-C5	-2.06	1.42	1.45
3	A	601	NAG	O5-C5	-2.01	1.42	1.45

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	601	NAG	O5-C5-C6	4.01	111.19	106.98
3	A	421	NAG	C2-N2-C7	3.64	127.80	123.39
3	A	421	NAG	C8-C7-N2	3.27	122.30	116.12
3	A	421	NAG	O5-C5-C6	3.14	110.27	106.98
3	A	421	NAG	C3-C2-N2	2.66	115.61	111.62

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	C	471	NAG	C1

There are no torsion outliers.

There are no ring outliers.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues

There are no chain breaks in this entry.

## 6 Fit of model and data ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	317/330 (96%)	-0.06	4 (1%) 74 82	52, 65, 88, 112	0
1	C	330/330 (100%)	-0.07	3 (0%) 81 88	49, 69, 104, 122	0
1	E	318/330 (96%)	0.26	17 (5%) 25 31	65, 94, 117, 135	0
2	B	173/173 (100%)	-0.05	2 (1%) 75 83	49, 60, 81, 106	0
2	D	172/173 (99%)	-0.10	0 100 100	48, 61, 81, 91	0
2	F	172/173 (99%)	-0.13	1 (0%) 86 91	52, 75, 100, 106	0
All	All	1482/1509 (98%)	-0.00	27 (1%) 65 74	48, 70, 108, 135	0

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	222	TRP	3.9
1	E	278	VAL	3.3
1	C	81	TYR	3.3
1	E	196	ILE	3.0
1	E	142	GLY	2.9

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
4	NAG	A	432	14/15	0.39	5.82	110,116,123,128	0
6	FUC	B	200	10/11	0.31	5.76	107,110,114,114	0
8	NAG	C	442	14/15	0.24	5.59	86,97,106,114	0
7	NAG	C	411	14/15	0.41	5.22	121,126,131,135	0
7	NAG	C	431	14/15	0.41	5.19	117,123,129,130	0
7	NAG	C	412	14/15	0.48	5.09	131,142,146,149	0
7	NAG	E	601	14/15	0.35	4.90	109,116,121,124	0
6	NAG	B	201	14/15	0.29	4.70	93,99,104,104	0
7	NAG	C	432	14/15	0.64	3.50	124,135,142,146	0
10	FUC	D	200	10/11	0.30	3.44	103,112,114,116	0
4	NAG	A	431	14/15	0.27	3.06	92,98,104,108	0
5	GAL	E	702	11/12	0.27	2.91	120,122,127,128	0
10	NAG	D	201	14/15	0.28	2.74	99,105,109,111	0
8	NAG	E	431	14/15	0.38	2.26	120,123,125,128	0
8	NAG	E	432	14/15	0.37	1.73	123,129,133,134	0
7	NAG	C	421	14/15	0.31	1.41	100,109,112,118	0
9	FUC	C	450	10/11	0.17	1.25	79,80,83,85	0
7	NAG	E	441	14/15	0.17	0.61	96,102,106,111	0
9	NAG	C	451	14/15	0.15	0.58	62,70,77,80	0
5	SIA	E	701	20/21	0.24	0.26	107,112,119,120	0
5	FUC	A	704	10/11	0.22	0.06	96,98,101,102	0
5	FUC	E	704	10/11	0.26	-0.12	124,127,131,131	0
5	SIA	C	701	20/21	0.17	-0.33	69,71,77,79	0
5	SIA	A	701	20/21	0.16	-0.47	65,69,75,76	0
7	NAG	C	601	14/15	0.17	-0.49	83,91,94,99	0
5	FUC	C	704	10/11	0.16	-0.82	94,95,98,98	0
5	GAL	A	702	11/12	0.16	-1.06	79,82,87,87	0
8	NAG	C	441	14/15	0.11	-1.60	69,74,77,83	0
5	NGS	C	703	19/19	0.12	-1.72	93,100,109,113	0
5	GAL	C	702	11/12	0.12	-2.74	81,84,89,89	0
4	BMA	A	433	11/12	0.39	-	133,140,147,148	0
9	BMA	C	453	11/12	0.33	-	112,120,130,134	0
7	NAG	E	442	14/15	0.38	-	116,120,122,123	0
4	MAN	A	437	11/12	0.41	-	146,151,159,160	0
7	NAG	C	422	14/15	0.32	-	120,127,131,135	0
5	NGS	A	703	19/19	0.19	-	90,96,107,111	0
7	NAG	C	602	14/15	0.27	-	94,109,113,117	0
7	NAG	E	602	14/15	0.37	-	118,128,131,134	0
8	BMA	E	433	11/12	0.29	-	135,138,143,145	0
6	NAG	B	202	14/15	0.38	-	99,107,112,114	0
9	NAG	C	452	14/15	0.16	-	84,88,97,102	0
9	MAN	C	457	11/12	0.35	-	137,141,146,146	0
9	MAN	C	454	11/12	0.51	-	117,124,131,136	0

Continued on next page...

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
8	BMA	C	443	11/12	0.28	-	120,124,130,132	0
5	NGS	E	703	19/19	0.22	-	131,135,141,145	0

## 6.4 Ligands

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	NAG	A	601	14/15	0.32	6.91	89,94,100,100	0
3	NAG	F	201	14/15	0.31	4.60	107,110,114,116	0
3	NAG	C	471	14/15	0.30	2.80	106,109,110,110	0
3	NAG	E	411	14/15	0.42	2.61	128,133,138,140	0
3	NAG	A	421	14/15	0.28	2.56	96,101,107,109	0
3	NAG	A	441	14/15	0.20	1.06	76,79,82,82	0

## 6.5 Other polymers

There are no such residues in this entry.